PDBj
Protein Data Bank Japan
http://pdbj.org/

January, 2014

Protein Data Bank Japan
NEWS LETTER vol. 15 no. 2

wwPDBAC meeting Report

The 10th wwPDB Advisory Committee meeting

The wwPDB (worldwide Protein Data Bank), of which the PDBj is one of the members, organizes the annual advisory Committee (wwPDBAC) meeting. This year, RCSB-PDB hosted the 10th meeting on September 27, 2013. The participants were Prof. Soichi Wakatsuki as a chair (SLAC-SSRL, Stanford), Prof. Helen M. Berman and Dr. Martha Quesada (RCSB-PDB, Rutgers Univ.), Dr. Gerard Kleywegt (PDBe, EBI), Prof. John L. Markley (BMRB, Univ. Wisconsin-Madison), Prof. Haruki Nakamura (PDBj, Osaka Univ.), Prof. Paul Adams (Lawrence Berkeley National Laboratory and UC Berkeley), Prof. Cynthia Wolberger (Johns Hopkins Univ.), Prof. Guy Montelione (Rutgers Univ.), Prof. Angela M. Gronenborn (Univ. of Pittsburgh), Prof. Andreas Engel (Case Western Reserve Univ.), Prof. Cynthia Wolberger (Johns Hopkins Univ.), Prof. Janet Thornton (EBI), Prof. Randy Read (Cambridge Univ.), Prof. Genji Kurisu (IPR, Osaka Univ.), Prof. Wah Chiu (Baylor College of Medicine), Prof. Edward Baker (Auckland Univ.) as a representative of IUCr, and Dr. R. Andrew Byrd (Center for Cancer Research, NCI) as a representative of ICMRBS. In addition, Prof. Jianping Ding (Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China) and Prof. Manju Bansal (Indian Institute of Science) attended the meeting as the associated member. (Fig. 1). After an overview was made by Prof. Helen M. Berman in RCSB-PDB, the progress of Common Deposition and Annotation (D & A) program was introduced and its current software system was shown by Dr. Martha Quesada and Jasmine Young, representing the team developing the Common D & A program. This system will soon work in 2014.

Fig. 1: Participants of the 10th wwPDBAC meeting

Next, as previously agreed by the wwPDBAC, the new PDB format, named as PDBx/mmCIF, was introduced by Dr. Gerard Kleywegt. A pipeline for creating a Validation Report for each crystal structure in PDB has been constructed, and several examples were shown. Validations for NMR and EM experimental data are also planned. The new formats for NMR experimental data, BMRB/XML and BMRB/RDF, have been developed by PDBj-BMRB team, and they were introduced by John L. Markley.

In summary, continuous efforts of the wwPDB members for the Common D & A program, the new PDBx/mmCIF format, and Validations were appreciated. More efforts are requested for announcement about those issues in Asian region. In addition, it was requested that the wwPDB should make itself clearly visible as the global organization.

Transition to the “new” PDB format

The “PDB format” means that fixed-column length, line-based flat file format to most users of the PDB. But the PDB format is more than 40 years old and it is unfortunately failing to meet the requirements today's structural biology imposes. For example, the number of atom that can be contained in a single entry is less than 100,000 so that some huge supramolecular structures are currently divided into several PDB entries. Also, today's structural biology cannot be done without referencing the results of other fields than structural biology so that functional, taxonomic, and other annotations are becoming as valuable as atomic coordinates themselves. But in the traditional PDB files it is sometimes difficult to parse such information without excessive “exception handling.” In order to address these issues, it was decided in 2011 wwPDB advisory committee that a new PDB format should be devised, and that new format turned out to be the PDBx/mmCIF format which is currently adopted as the official archiving format of the wwPDB. Starting from 2014, large structures that cannot fit the legacy PDB format will be released only in mmCIF and PDBML formats (some large structures are already tentatively being released as such: see http://pdbj.org/mine-search?query=tag%3Alarge_structures).

-Transition to the “new” PDB format

So, what's good about PDBx/mmCIF? First of all, the grammar of PDBx/mmCIF is rigorously defined so that no “exception handling” is necessary as far as the syntax is concerned. Second, the categories and terms in the format are precisely defined in the PDBx dictionary. By looking up the dictionary, we can learn which category describes what kind of data, as well as the relationships between different categories. The syntax of the PDBx dictionary itself is formally defined so that we can parse the dictionary to generate a program than manipulates PDBx/mmCIF files, for example. At present, there are over 300 categories defined in the dictionary, among them are references to external data resources, which facilitates integration of structural data with other biological and/or chemical annotations. Third, compared to PDBML (a “direct” translation of mmCIF into XML), PDBx/mmCIF files are (arguably) easy to read both by humans and by machines: tags describing category and category items are simple enough to not interfere human eyes.

Given the above benefits, what are obstacles for adopting the PDBx/mmCIF format? Apparently, many programs do not support the format yet. However, some widely used applications already support the PDBx/mmCIF format (e.g., Jmol, OpenRasMol, Chimera, CCP4, Phenix). Programmers may feel reluctant to modify their codes. But there are PDBx/mmCIF parsers already available in some major programming languages such as C/C++, Python, Java, and Perl. By using such libraries, the amount of codes to be modified may be greatly reduced.

A comprehensive list of PDBx/mmCIF-related information can be found at *http://mmcif.wwpdb.org*.
In addition, some learning material is also available in Japanese: *http://pdbj.org/info/new-format*

Please do not hesitate to ask us your questions about this big issue of the format transition from "Contact us" page in our Web site (http://pdbj.org/)

Towards database integration

PDBj has been developing a new system towards database integration for wide biological research. In particular, several tools have been created for function annotation from genome sequence through protein tertiary structures. The followings are our latest tools and databases.

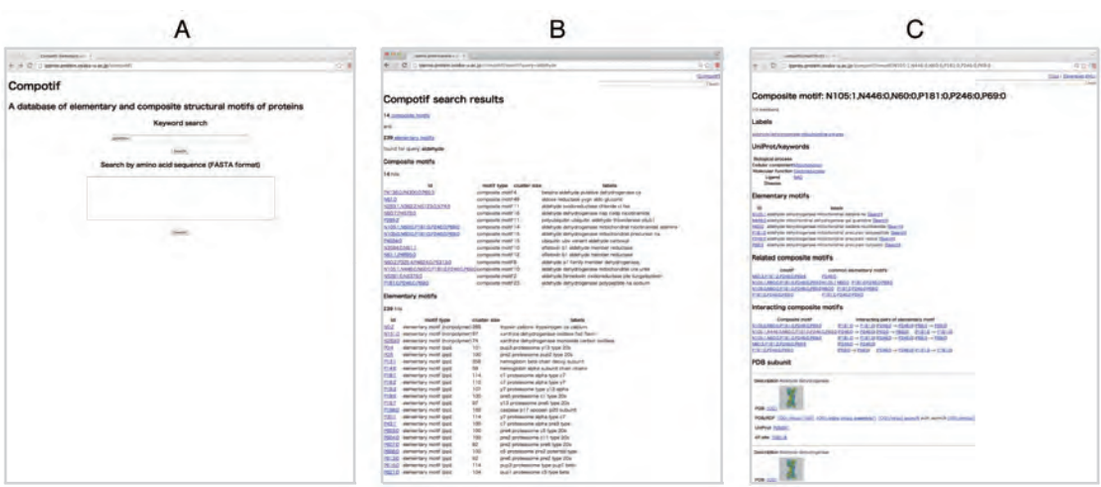
1. Compotif: A database of elementary and composite structural motifs of proteins

http://ipproo.protein.osaka-u.ac.jp/compotif/

Compotif is a database that compiles structural motifs of interaction interfaces in proteins. There are three types of interactions considered that include those with small compounds, proteins, as well as nucleic acids (DNA/RNA). The basic information contained in Compotif are following [1]:
(1) Elementary motifs classified according structural similarities of single interfaces.
(2) Composite motifs that combine multiple elementary motifs co-occurring in single subunits.
(3) Relationships between different composite motifs (sharing same elementary motifs) and between composite and elementary motifs.

The elementary motifs have been identified based on exhaustive structure comparisons using the GIRAF program [2] provided by the PDBj. The user interface of Compotif is very simple: input keywords or amino acid sequence (Figure 2-A), and follow the links of elementary and composite motifs in the hit list (Figure 2-B). Each page of motifs contains various links to other motifs and external databases including PDBj, UniProt and eF-site (Figure 2-C).

Fig. 2: Search for the structural motifs by Compotif



2. eF-patch: A database for molecular surfaces of protein ligand binding sites

http://ef-patch.hgc.jp

Many proteins have biological functions such as regulation of biological activity and signal transduction by interacting with other small molecular compounds (ligand). In order for proteins to selectively recognize a specific ligand, it is considered that ligand-binding sites in proteins can have characteristic pocket shapes and electrostatic potentials. Recently, we have exhaustively compared and clustered local molecular surfaces for the atomic coordinates of ligand-binding pockets obtained from protein-ligand complex structures in PDB. Such local surfaces (patch) were extracted from molecular surfaces of proteins stored at eF-site [3]. Consequently, we have found that some similar patches are conserved in distantly related proteins and even in proteins with cross-fold similarity, and proteins with highly similar patches have potential to be involved in similar biological processes [4]. Then, we have compiled a database of the clustering results of patches and integrated distributed life-science databases for information of protein sequence (UniProt), structure (PDB/RDF, PDBj, eF-site, SCOP and CATH) and function (keywords in UniProtKB) over the Internet using uniform resource identifiers (URI). Therefore, this integrated DB system, named as eF-patch (http://ef-patch.hgc.jp), enables rational inference of protein functions from different viewpoints of information from protein sequence and structure. In addition, each data entry of eF-patch can be accessed from search results of eF-seek [5], which is a web server to detect similar patches for uploaded atomic coordinates of proteins, so that it enables to perform from identification of binding sites to inference of functions in proteins. Furthermore, eF-patch allows researchers all over the world to correct and reuse our data for their purposes by including micro-date for semantic markup in web pages for each entry in eF-patch, in order to contribute to the development of life sciences.

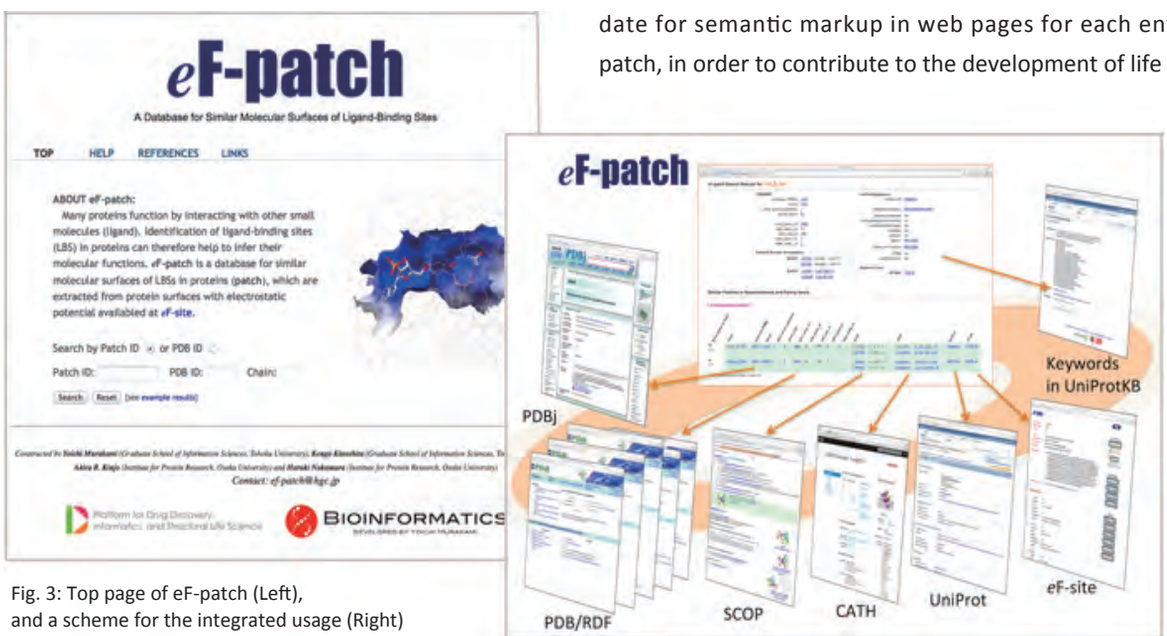


Fig. 3: Top page of eF-patch (Left), and a scheme for the integrated usage (Right)

3. SFAS: An improved Sequence to Function Annotation Service pipeline

http://sysimm.ifrec.osaka-u.ac.jp/sfas3/

SFAS is a threading meta server that interfaces with the 3D modeling package Spanner (http://sysimm.ifrec.osaka-u.ac.jp/spanner/) and the functional annotation tool SeSAW (http://sysimm.ifrec.osaka-u.ac.jp/SeSAW/). The design of SFAS is meant to be intuitive and to facilitate manual interaction with the threading programs based on 2D information (intrinsic disorder prediction and secondary structure). Intrinsically disordered domains can be sent to IDD navigator (http://sysimm.ifrec.osaka-u.ac.jp/disorder/beta.php) for sequence composition-based function prediction. Based on feedback from our beta-version of SFAS (http://sysimm.ifrec.osaka-u.ac.jp/sfas2/) we have developed an improved version, which is now located at the temporary URL (http://sysimm.ifrec.osaka-u.ac.jp/sfas3/). There are two main areas where the new pipeline has been improved.

The first is in the back-end threading programs. The threading programs are third-party code and include a large amount of data such as processed PDB entries, which can easily be aligned to the query sequence of interest. In the former version of SFAS we attempted to produce these data files ourselves, using tools provided by the threading program developers ; however, this

resulted in inconsistent alignment results. In the new version, we have decided use the provided data files. The resulting alignments are then mapped to our local copy of PDB entries.

The second improvement is in the interface (Fig. 4). The 2D prediction page (intrinsic disorder and secondary structure) is now much easier to see and manipulate. The 3D prediction page now has more choices for 3D viewing and the alignments are shown in an expanded window.

This new SFAS server is scheduled to be complete by mid January of 2014.

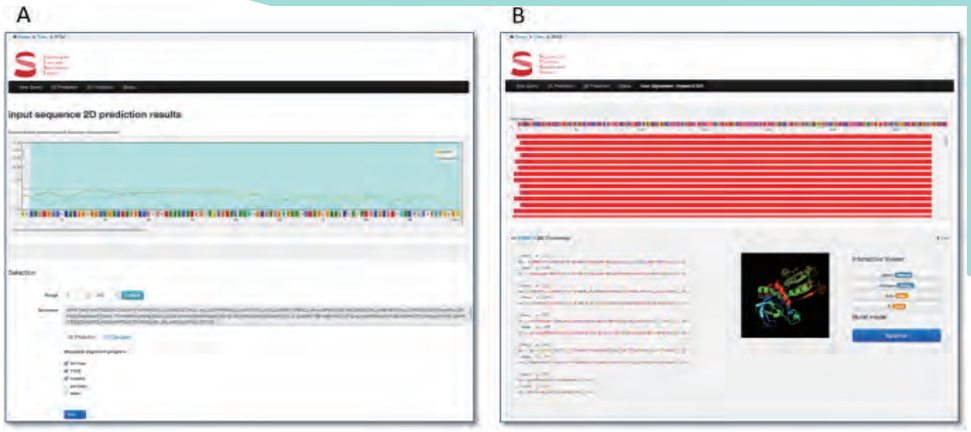


Fig. 4: Screenshots of the new SFAS server.
Left: The 2D prediction result window, from which sub-sequences can be selected and exported to either IDD Navigator or to 3D threading servers. Right: The 3D prediction result window showing an overview of alignments, and a particular pairwise query-template alignment with several viewer choices as well as the option for exporting to Spanner.

References

1. Kinjo AR, Nakamura H. Composite structural motifs of binding sites for delineating biological functions of proteins. PLoS One 7:e31437 (2012)

2. Kinjo AR, Nakamura H. GIRAF: a method for fast search and flexible alignment of ligand binding interfaces in proteins at atomic resolution. Biophysics 8:79-94 (2012)

3. Kinoshita K, Nakamura H. eF-site and PDBjViewer: database and viewer for protein functional sites. Bioinformatics 20:1329–1330 (2004)

4. Murakami Y, Kinoshita K, Kinjo AR, Nakamura H. Exhaustive comparison and classification of ligand-binding surfaces in proteins. Protein Sciences 23:1379-1391 (2013)

5. Kinoshita K, Nakamura H. Identification of the ligand binding sites on the molecular surface of proteins. Protein Science 14:711–718 (2005)

Data Growth

The statistics data is also available at the wwPDB web page (http://wwpdb.org/stats.html).

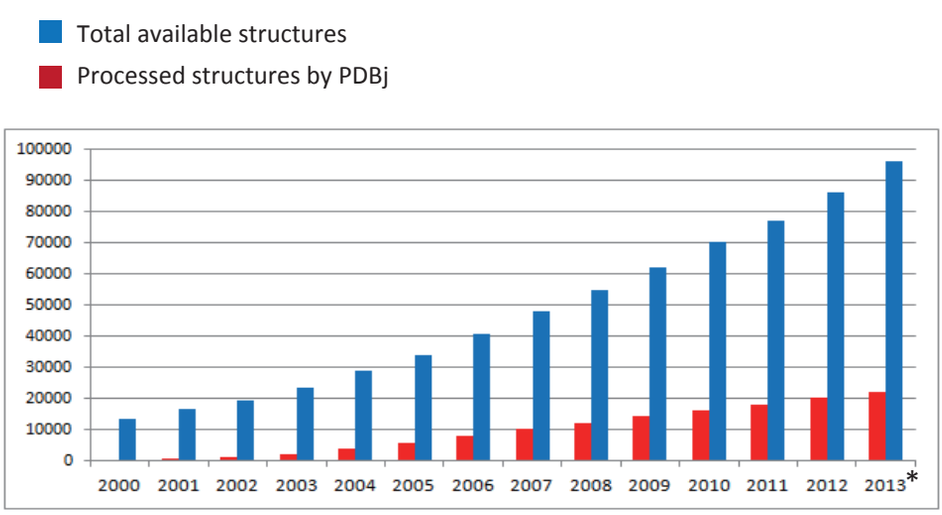


Fig.5

*96798, as of January 8, 2014

Third wwPDB Working Group on Theoretical Model Validation

OCT 21-22, USA

The Third Working Group on Theoretical Model Validation was held from October 21-22, 2013 at Rutgers University, NJ. As the title implies, this was the third in a series of modeling workshops. The distinguishing feature of this particular workshop was the emphasis that concrete actions should be taken to provide a resource to house structural models that are published in scientific journals. To this end, small group discussions were held to determine the requirements for such a resource.

wwPDB workshop on particular a united restraints format

NOV 17-19, UK

wwPDB workshop on a united restraints format, of which an international protein structure data bank (wwPDB) organizes, was held in EMBL-EBI (European Molecular Biology Laboratory European Bioinformatics Institute), at Hinxton, United Kingdom from November 17-19, 2013.

Fig. 6: Participants of the wwPDB workshop on a united restraints format

Dr. Naohiro Kobayashi from the PDBj-BMRB group in Osaka University participated, and it had a discussion about establishment of the compatible format for the restraint information at the time of NMR structure deposition to PDB.

A number of researchers (Fig. 6) participated in the meeting, including representative members of database of NMR structure and experimental data in addition to the developer of NMR structural analysis tools, such as Cyana, UNIO, ARIA and XPLOR-NIH. The discussion was performed from various angles. On Nov. 17th, NMR Validation Task Force Meeting preceded to be held at the same place, and the practical validation methods of NMR structure and experimental data were argued.
More detailed arguments on the united restraints format for the NMR structural determination, which were aimed to adopt validation method, were performed in 18-19th, then the guidelines for establishing the new format were determined.

wwPDB Workshop on mmCIF/PDBx for Programmers

NOV 20-21, UK

wwPDB Workshop on PDBx/mmCIF for Programmers organized by the internationalprotein structure data bank (wwPDB) was held from November 20-21, 2013 in EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) at Hinxton, United Kingdom (Fig. 7).

Takahiro Kudo, a PDBj staff, took part in the workshop and worked on how to handle technically the mmCIF format data. The remarkable feature of this event was that all attendees could practice various tools with C++, Python, and Java, then we got a great deal of information for using mmCIF practically. On the basis of those, PDBj will provide helpful information about mmCIF for our users.

Fig. 7: Participants of the wwPDB Workshop on PDBx/mmCIF for Programmers

AsCA'13: Asian Crystallographic Association Meeting in 2013

DEC 7-10, HongKong, HKUST

In the Asian Crystallographic Association Meeting (AsCA) held in Hong Kong from December 7-10, 2013, Prof. Genji Kurisu (IPR, Osaka University), who is an Advisory Committee member of PDBj, had an oral presentation entitled “Modifications to the Protein Data Bank” . Some modifications planned to start early 2014, such as new PDBx format, new Data deposition and Validation report, were announced to the biological crystallographers from Asia and Oceania.

The local organizing committee of AsCA’ 13 kindly gave us an opportunity to present a 25 minutes talk in the beginning of biological sessions. After the talk, some comments and questions were came up from the audience, mainly concerning about the ambiguous stereo chemical assignment of ligants at low resolution. It must be important for PDBj to keep in touch with AsCA, as one of the scientific societies closely related to PDB activities.

Event Report

Activities done for introducing PDBj and our newly developed web services.

- Luncheon Seminar at 13th Annual Meeting of the Protein Science Society of Japan (JUNE 12, Torigin Bunka Kaikan, Tottori-city)
- Workshop on PDBj web & databeses (AUG 23, Nakanoshima Center, Osaka-city)
- Luncheon Seminar at the Annual Meeting 2013 of Crystallographic Society of Japan (OCT 13, Kumamoto-city)
- Luncheon Seminar at 51th Annual Meeting of the Biophysical Society of Japan (OCT 28, Kyoto-city)
- Exbition Booth at Science Agora 2013 -Science event for the public, supported by Japan Science and Technology Agency (NOV 9-10, Odaiba, Tokyo)
- Exbition Booth and Presentation at 36th Annual Meeting of the Molecular Biology Society of Japan (DEC 3-5, Kobe-city, Hyogo)

Fig.8: Luncheon Seminar at the Annual Meeting 2013 of Crystallographic Society of Japan

Fig.9: Science Agora 2013

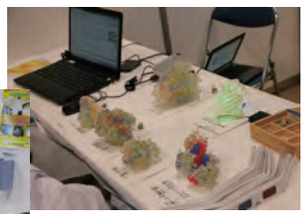


Fig.10: 3D Structure models for exbition at Science Agora

Fig.11: Presentation at 36th Annual Meeting of the Molecular Biology Society of Japan

Staff

Head Nakamura, Haruki, Ph. D. (Prof., IPR, Osaka University)

Group for PDB Database Curation

Nakagawa, Atsushi, Ph. D. (Prof., IPR, Osaka Univ.)
Matsuda, Makoto, Ph. D. (IPR, Osaka Univ.)
Igarashi, Reiko (IPR, Osaka Univ.)
Kengaku, Yumiko (IPR, Osaka Univ.)
Cho, Hasumi, Ph. D. (IPR, Osaka Univ.)
Ikegawa, Yasuyo (IPR, Osaka Univ.)
Sato, Junko (IPR, Osaka Univ.)

Group for BMRB

Fujiwara, Toshimichi, Ph. D. (Prof., IPR, Osaka Univ.)
Akutsu, Hideo, Ph. D. (Guest Prof., IPR, Osaka Univ.)
Kojima, Chojiro, Ph. D. (IPR, Osaka Univ.)
Kobayashi, Naohiro, Ph. D. (IPR, Osaka Univ.)
Iwata, Takeshi (IPR, Osaka Univ.)
Takahashi, Ami (IPR, Osaka Univ.)
Yokochi, Masashi (IPR, Osaka Univ.)

Group for Development of new tools and services

Kinjo, Akira R., Ph. D. (IPR, Osaka Univ.)
Iwasaki, Kenji, Ph. D. (IPR, Osaka Univ.)
Suzuki, Hirofumi, Ph. D. (IPR, Osaka Univ.)
Yamashita, Reiko (IPR, Osaka Univ.)
Kudou, Takahiro (IPR, Osaka Univ.)
Nishikawa, Ken, Ph. D. (Guest Prof., IPR, Osaka Univ.)
Bekker, Gert-Jan (IPR, Osaka Univ.)

Collaboratory Researchers

Wako, Hiroshi, Ph. D. (Prof., Waseda Univ.) for Pro Mode
Ito, Nobutoshi, Ph. D. (Prof., Tokyo Medical and Dental Univ.)
Kinoshita, Kengo, Ph.D. (Prof., Tohoku Univ.) for e F-site
Standley, Daron, Ph. D. (IFReC, Osaka Univ.)
for SeqNavi, StructNavi, SeSAW, and ASH
Katoh, Kazutaka, Ph. D. (IFReC, Osaka Univ.) for MAFFTash

Secretary Haruki, Nahoko (IPR, Osaka Univ.)

Contact

Protein Data Bank Japan

http://pdbj.org/

Research Center for State-of-the-Art Functional Protein Analysis,
Institute for Protein Research, Osaka University
3-2 Yamadaoka, Suita, Osaka 565-0871, JAPAN

PDBj Office TEL: +81-6-6879-4311 FAX: +81-6-6879-8636
PDBj Deposition Office TEL: +81-6-6879-8634 FAX: +81-6-6879-8636